



King's Research Portal

DOI:

[10.1017/thg.2018.9](https://doi.org/10.1017/thg.2018.9)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Jarrar, Z. A., Ward, K. J., Mangino, M., Cherkas, L. F., Gill, R., Gillham-Naseny, I., Yarand, D., Hart, D., Hammond, C. J., & Spector, T. D. (2018). Definitive zygoty scores in the peas in the pod questionnaire is a sensitive and accurate assessment of the zygoty of adult twins. *TWIN RESEARCH AND HUMAN GENETICS*, 21(2), 146-154. <https://doi.org/10.1017/thg.2018.9>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Zygoty assessment using peas in the pod questionnaire

Definitive zygoty scores in the peas in the pod questionnaire is a sensitive and accurate assessment
of the zygoty of adult twins

Zakariya A. Jarrar[^], Kirsten J. Ward[^], Massimo Mangino, Lynn F. Cherkas, Raj Gill, Irina Gillham-
Nasanya, Darioush Yarand, Deborah Hart, Christopher.J. Hammond*, Tim D. Spector*

Department of Twin Research and Genetic Epidemiology, King's College London, Block D, South
Wing, 3rd and 4th Floor, St. Thomas' Hospital Campus, Westminster Bridge Road, London, SE1 7EH,
United Kingdom.

[^]Joint first author * Joint last author

Corresponding author – T.D. Spector, tim.spector@kcl.ac.uk

Abstract:

Twin researchers face the challenge of accurately determining the zygoty of twins for research. As part of the annual questionnaire between 1999 and 2006, 8,307 twins from the TwinsUK registry were asked to complete 5 questions (independently from their co-twin) to ascertain their self-perceived zygoty during childhood on up to 5 separate occasions. This questionnaire is known as the “peas in the pod” questionnaire (PPQ) but there is little evidence of its validation. Answers were scored and classified as monozygoty (MZ), dizygoty (DZ) or unknown zygoty (UZ) and were compared with 4,484 twins with genotyping data who had not been selected for zygoty. Of these, 3,859 individuals (46.5% of those that had a zygoty from PPQ) had zygoty classified by both the PPQ and genotyping.

Of the 708 individual twins whose answers meant that they were consistently classed as MZ in the PPQ, 683 (96.5%) were MZ within the genotype data. Of the 945 individual twins consistently classed as DZ within questionnaire, 936 (99.0%) were DZ in the genotype data. Where both twins scored MZ consistently across questionnaires, 99.6% were MZ on genotyping; 99.7% were DZ on genotyping if both twins consistently scored DZ. However, at initial questionnaire 88.6% of those scoring as MZ were genotypically MZ and 98.7% DZ. For twin pairs where both scored UZ, 94.7% were DZ. Using the PPQ on a single occasion provided a definitive classification of whether the twin was monozygoty or dizygoty with an overall accuracy of 86.9%, increasing to 97.9% when there was a consistent classification of zygoty across multiple questionnaires.

This study has shown that the “peas in the pod” questionnaire is an excellent proxy indicator of zygoty in the absence of genotyping information.

Introduction:

A common misunderstanding, until relatively recently, was that the presence of two placentae at birth must mean that twins are non-identical or dizygotic (DZ) whilst one placenta meant that the twins are identical or monozygotic (MZ)(Ooki et al., 2004). Opposite sex twins are DZ, but there are issues arising when using this methodology to determine the zygoty of same sex twins. This meant that between 20-25% of MZ twins with two separate placentae were misclassified as DZ at birth, with approximately 9% of DZ twins with placental fusing wrongly classed as MZ (Ooki et al., 2004). As a result a number of twin pairs have grown to adulthood understanding that they are identical or non-identical when they are not. This misclassification of zygoty at birth has implications on not only the medical treatment of twins during gestation, after their birth and in later life, but also has personal implications on the twins who may grow up questioning their zygoty.

Within the scientific community, an essential component of any twin registry is the knowledge of the zygoty of the twins registered. There are a number of different methods of determining the zygoty in adult twins. These include self-report methods, for example asking the twins whether they are identical or non-identical, asking twins questions on how similar they are or were during childhood and growing up (Cederlöf et al. 1961, Sarna et al. 1978, Magnus et al. 1983, Bønnelykke et al. 1988, Ooki et al. 1989, Song 2010) or asking the parents to report on the similarity of twins (Peeters 1998).

Determination of zygoty by self-report raises concerns that misclassification may have occurred at birth by an authoritative medical professional. More reliable methods include testing by blood group (Cederlöf et al. 1961, Kasriel & Eaves 1976, Lykken 1978) or using genetic tests (Song 2010) but these may not necessarily be fast or cost-effective. Song *et al* used 16 short tandem repeat genetic markers and demonstrated that using questions about similarities during childhood provided a sensitivity of 98.8% for MZ twin pairs and 88.9% for DZ twin pairs.

Genotypic data is the “gold” standard for determining the zygoty of twins, but as this is costly and time consuming when testing a large scale study at one time, alternative methods of accessing zygoty are necessary, and as a result latent class analysis has been used with good success (Eaves et al. 1993, Heath et al. 2003).

The TwinsUK Adult Twin Registry started in 1992 (Spector & Williams 2006, Moayyeri et al. 2013, Moayyeri et al. 2013) and as with all twin registries, has faced the challenge of determining the zygoty of the twins. The aims of this study were to compare the accuracy of zygoty determined by a “peas in the pod” similarity questionnaire (the PPQ) for both single twins and twin pairs with the zygoty determined by genotype data, the “gold standard”, and to examine the consistency of responses when the PPQ is administered annually compared to initial self-report and PPQ.

Materials and Method:

Using a short similarity questionnaire used by the Australian Twin Registry as a basis (unpublished, and known as the peas in the pod questionnaire or PPQ), the TwinsUK registry has adapted this to be more specific to the cohort. Between 1999 and 2006, the TwinsUK registry asked same sex twins (aged 18 to 89 years of age) the PPQ (see S1) as part of their annual questionnaire. The accompanying instructions asked for the questionnaire to be completed by each twin separately to ensure that the twins were not influenced in the answers they gave by their co-twin. Written, informed consent was obtained from participants of the study and all procedures contributing to this work comply with the ethical standards in the Helsinki Declaration of 1975, as revised in 2008.

Self-reported zygosity

Upon registration with the TwinsUK registry, each twin was asked to report what they believed their zygosity to be. A total of 13,291 twins had provided a self-reported zygosity (age 18 to 84, mean age of 58; 10,796 (83%) were female, 2,226 were male and 269 had no gender assigned to them).

Peas in the Pod Questionnaire (PPQ)

The PPQ is a 5-item questionnaire on the degree of similarity between twins (Supplementary Material 1). It asks four questions based on whether at school-age people at school, parents, close-friends, or strangers had difficulty telling twins apart (0=yes, 1=don't know, 2=no), and a fifth question whether the twins would be described during childhood as alike as two peas in a pod (0 points), as alike as ordinary siblings (2 points) or they didn't know (1 point). Scores between 0 and 4 were classed as MZ; scores between 8 and 10 were classed as DZ and anything in between was scored as unknown zygosity (UZ), with each twin's PPQ scored separately.

A total of 8,307 twins answered a PPQ zygosity (age 18 to 87, mean age of 51; 7,287(88%) were female and 1,020 were male) between 1999 and 2006.

Initial analysis looked at the zygoty provided from the first questionnaire asked and then the individual twins were matched to their co-twin. The second stage of analysis looked at all data collected from the PPQ between 1999 and 2006. The PPQ was scored separately for each questionnaire on 5 separate occasions so that the scores for each questionnaire and therefore the zygoty of the twin for each questionnaire were independent of the scores and zygoty from the other PPQ questionnaires.

To create an overall zygoty over time, individuals who consistently had the same zygoty for each of the questionnaires was scored as that zygoty (consistent answers of MZ meant an overall zygoty of MZ, consistent answers of DZ meant an overall zygoty of DZ and consistent answers of UZ meant an overall zygoty of UZ) and individuals who zygoty was inconsistent over time e.g. MZ in one questionnaire and DZ in another were scored as UZ.

The overall zygoty on an individual and paired basis was determined from all of the 5 PPQ questionnaires. If both of the twins had the same overall zygoty, they were scored as this zygoty (i.e. if both had an overall zygoty of DZ then the pair was scored as DZ) and if the twins disagreed within the pair, this was noted so that it would be possible to see which zygoty was correct after comparison with the zygoty determined via genotyping.

To determine the weight of each question within the PPQ, the final stage was to look at each individual question within the PPQ and score the answer to each question as MZ (if the answer scored 0), UZ (if the answer scored 1) and DZ (if the answer scored 2).

To assess the sensitivity of the PPQ, a standard true positive rate calculation was used and to assess the specificity of the PPQ, a standard true negative rate calculation was used.

Genotype data

As MZ twins share 100% of their segregating genes, only one MZ twin of a pair was routinely genotyped for GWAS, whereas both members of a DZ or UZ pair were genotyped.

[Insert Table 1 here]

TwinsUK samples were genotyped with the Infinium 317K and 610K assays (Illumina, San Diego, USA) at two different centres, the Wellcome Trust Sanger Institute and the Center for Inherited Diseases Research (USA) respectively. The normalized intensity data was pooled and the genotypes were called on the basis of the Illumina algorithm. No calls were assigned if the most likely call was less than a posterior probability of 0.95. Validation of pooling was done by visual inspection of 100 random, shared single nucleotide polymorphisms (SNPs) for overt batch effects; none were observed. We excluded SNPs that had a call rate <97% (SNPs with minor allele frequency (MAF) $\geq 5\%$) or <99% (for $1\% \leq \text{MAF} < 5\%$), Hardy-Weinberg p values $< 10^{-6}$ and MAFs $< 1\%$. We, also removed subjects where genotyping failed for >2% of SNPs. The overall genotyping efficiency of the genome wide association (GWA) was 98.7%.

A total of 4,484 twins had zygoty determined via genotyping (age 20 to 90 mean age of 61; 4,136 (92%) were female and 348 were male).

We computed identity-by-descent (IBD) estimates for all available pair of twins (n=4,484 individuals) using the Plink (Purcell S 2007) option *--genome* on a set of 9357 SNPs (not in linkage disequilibrium, with a minor allele frequency > 20% and overlapping among all Illumina platforms available). For ambiguous cases, IBDs were subsequently recalculated by using all the SNPs available. We defined as MZ those twin pairs with a pi hat value < 0.9 and DZ the twin pairs with a pi hat value ranging between 0.4 and 0.6.

The results from the PPQ

The PPQ has been used by the TwinsUK registry since the start of the registry in 1992, but for the purpose of this study, we concentrated on the data that was available over a 7 year period of time (between 1999 and 2006), where the PPQ was asked within the annual questionnaire on 5 separate occasions. It was possible to determine the zygosity for 8,307 individuals who had answered the PPQ on at least one occasion.

Results:

Since the start of the TwinsUK registry in 1992, 13,291 twins have provided a self-reported zygosity (including 6,644 complete twin pairs). Of these, 6,129 (46.1%) reported that they were MZ, 6,359 (47.8%) reported that they were DZ and 803 (6.1%) reported that they were UZ.

Zygosity from individual twins

Looking at the first PPQ answered by each twin separately, 4,038 (48.6%) individuals had scores that indicated that they were MZ, 1,645 (19.8%) provided answers that indicated that they were DZ and 2,624 (31.5%) were graded as UZ as they scored 5-7 on the PPQ (see figure 1).

Examining repeated PPQ's, 3,562 (42.9%) of the twins consistently scored as MZ and 1,536 (18.5%) of the twins consistently were rated as DZ on PPQ scores. The remaining 3,209 (38.6%) of twins' answers did not always result in the same grade (MZ, DZ, UZ) over time (n=3,062, 36.9%) or were scored consistently as UZ within the PPQ (n=147, 1.7%) see figure 1.

[Insert Figure 1 here]

Zygosity within twin pairs

From the 8,307 individuals who had answered the PPQ, there were 3,697 complete twin pairs (7,394 individuals). Of these, 1,387 pairs (37.5%) both scored that they were MZ, 480 pairs (13.0%) both scored that they were DZ and 1,150 (31.1%) both scored that they were UZ. The remaining 680 (18.4%) pairs could not agree on their zygosity. Of these 10 (1.5%) had one twin scoring as MZ and

the other twin scoring as DZ, 379 (55.7%) where one twin scored as MZ and the other twin as UZ and 291 (42.8%) where one twin scored as DZ and the other twin scored as UZ.

Comparison of zygosity from the PPQ and genotyping data

Of the 4,484 individuals from the TwinsUK registry with zygosity determined by genotype data, 994 (23.9%) were MZ and 3,162 (76.1%) were DZ. When the zygosity obtained from the PPQ was matched with the “true” zygosity from the genotyping data, there were 3,859 twins with zygosity for both methodologies which included 1,806 complete twin pairs (see figure 1).

Zygosity from individual twins

The zygosity from the first PPQ was matched with the genotyping data, 943 twins answered that they were MZ in the PPQ, of which 735 (77.9%) were MZ in the genotyping data. 1,101 answered that they were DZ, of which 1071 (97.3%) were DZ in the genotyping data. From the 1,811 twins that were UZ from their first PPQ, 1665 (91.9%) are DZ from the genotyping data (see figure 3). Using the zygosity obtained from the first PPQ proved to have 96.1% sensitivity and 83.7% specificity.

[Insert Figure 2 here]

Comparing the overall zygosity from the PPQ and genotype data there were 3,859 individuals that had zygosity determined by both methodologies (providing coverage of 46.5% of the PPQ zygosity results) (see table 2). Using the overall zygosity obtained from all of the PPQ's proved to have 98.6% sensitivity and 97.4% specificity.

[Insert Table 2 here]

Seven hundred and eight individual twins consistently indicated that they were MZ in the PPQ, of which 683 (96.5%) were MZ in the genotype data. Nine hundred and forty five twins consistently indicated that they were DZ in the PPQ, of which 936 individual twins (99.0%) were DZ in the

genotype data. From the 2,206 twins that were classed as UZ from the PPQ, 1,987 (90.1%) were DZ in the genotype data.

The first time that the PPQ is completed by an individual twin and compared to the genotyping data, there is 77.9% accuracy if the score is MZ, 97.3% accuracy if the score is DZ and when the score indicates a UZ, there is a 91.9% chance that they are in fact a DZ twin. This accuracy increases to 88.6% for MZ and 98.7% for DZ when both twins' scores indicate that they are MZ or DZ respectively.

Zygosity within twin pairs

Taking the answer from both twins from the first PPQ that they had answered showed that there were 343 pairs that both agreed that they were MZ, of which 304 (88.6%) were MZ in the genotype data. From the 390 pairs that both answered that they were DZ, 385 (98.7%) were DZ in the genotyping data (see figure 2).

[Insert Figure 3 here]

For twin pairs where it was not possible to determine the zygosity, 703 had both scored as UZ and 666 (94.7%) of these were DZ from the genotype data and 370 did not agree on their zygosity, of which 296 (80%) were DZ in the genotype data.

Looking at the overall PPQ scores from all questionnaires, there were 274 twin pairs where both scored as MZ consistently, of which 273 (99.6%) were MZ in the genotype data. From the 312 twin pairs where both twins consistently scored as DZ within the PPQ, 311 (99.7%) were DZ in the genotype data. There were 901 twin pairs where it was not possible to ascertain the zygosity of either twin from the PPQ (both twins scored UZ, 5-7 points). From these UZ pairs, 837 (92.9%) of the twins were DZ in the genotype data (see table 3).

[Insert Table 3 here]

There are a number of twin pairs, where only one twin had scored as a consistent zygoty within the PPQ over time. When one twin consistently scored as MZ in the PPQ and the other twin as UZ, comparison with the genotype data showed that the MZ twin score was correct on 87.8% occasions. When one twin consistently scored as DZ in the PPQ over time and the other twin as UZ, comparison with the genotype data showed that the DZ twin score was correct on 99.1% occasions.

Weighting of the individual questions within the PPQ

The answer for each question within the questionnaire was scored as MZ, DZ or UZ and compared with the result on zygoty obtained from the genotyping.

For determining both MZ and DZ twins (see figure 4), the most accurate question is question “e) In childhood, which of the following would best describe you and your twin”. This had an accuracy of 92.5% for MZ twins and an accuracy of 97.2% for DZ twins when compared to the genotyping data. The least accurate question for MZ twins was question “a) At school, did people have trouble telling you apart?” with an accuracy of 48.5% and for DZ twins question “b) Were your parents able to tell you apart?” with an accuracy of 80.3%.

These results suggest that it is possible to ascertain the zygoty of both MZ and DZ twins using just one question “e) In childhood, which of the following would best describe you and your twin” with the possible answers being “As alike as two peas in a pod”, “Ordinary sibling likeness (Like sisters or brothers)” and “I don’t know” (see Supplementary Material 1).

[Insert Figure 4 here]

Comparison between self-reported zygoty, zygoty from PPQ and genotyping data

Eighty-eight percent of the twins self-reported that they were MZ at registration were also MZ within the first PPQ that they answered (see figure 5). Eighty two percent of those that self-reported MZ’s remained MZ for the overall zygoty and 95.9% of those that self-reported as MZ remained as

MZ within the genotyping data. However, reporting that they were DZ at registration did not appear to be as consistently accurate as 36.3% were DZ within the first PPQ that they answered; 34.3% were DZ with the overall zygoty from all of the PPQ's that they answered. However, 94.4% of those that had self-reported as DZ were also DZ within the genotyping data.

[Insert Figure 5 here]

Discussion:

It may well be that the historical misclassification of zygoty according to number of placentae explains the discrepancies between the self-report zygoty and genotyping results, particularly in our cohort of twins born before human genomic testing was practical. 95% of individuals who self-reported as MZ and 94.4% as DZ had their zygoty confirmed by genotyping. In addition to self-report data, the TwinsUK registry has used the PPQ as an initial indicator of zygoty, and in the main now has the luxury of genetic techniques to confirm the zygoty of a twin pair. *Traditionally, genetic determination of zygoty has been seen as costly, particularly in large-scale research studies. However, with the reduction in costs of genotyping in recent years, it is not expensive when compared to the pregnancy and childbirth costs of twin pregnancies. We would certainly advocate routine genotypic testing of twins at birth, to allow families to have definitive zygoty ascertained.*

Interestingly, although the PPQ was designed to improve classification of zygoty from self-report, the score obtained from the first time that the PPQ is answered appears to be less predictive of true zygoty - only 77% who were scored from PPQ as MZ were confirmed as MZ from the genotyping data, compared to 97% genetically confirmed as DZ from those who were scored as DZ.

Indecision about zygoty in questionnaire studies has previously been suggestive that twins are likely to be dizygotic (Kasriel et al. 1976). Our results are similar which show that 92.9% of twins whose answers scored as unknown zygoty on multiple questionnaires were found to be dizygotic. This percentage is higher than a previous study where both twins answered as unknown zygoty (Song 2010) and could be explained by the fact that participants in our study were asked to complete multiple questionnaires over which may have increased the accuracy of the results. Eighty-eight per cent of twin pairs who answered differently to each other were found to be dizygotic. As the questionnaire was asked on several occasions, participants in our study were instructed not to confer with their fellow twin about their perceived zygoty so as not to influence their answers, and we did not ask them to come to any pair-wise decisions on zygoty.

There is variation in the literature about the accuracy of questionnaire-based zygoty. In agreement with our study, some studies confirm DZ more frequently than MZ on questionnaire (Peeters1998). However, other studies confirm MZ more than DZ (Cederlöf 1961). Whilst questionnaires used in different studies contain similar components, there is no standardized questionnaire used between studies, which may account for some of the differences. Also, older studies used techniques such as blood group to determine zygoty, which may be less accurate than modern genotyping technologies which have near 100% accuracy (Chen 2010). As discussed in the introduction, it may be that our particular twin population, on average born in the 1950's, was misinformed about zygoty due to misinformation from midwives/doctors based on numbers of placentas. Finally, there may be a bias that often only one of the pairs who were consistently MZ on PPQ was genotyped, reducing the numbers of MZ pairs who would likely have been confirmed by genotyping.

The single question "In childhood, which of the following would best describe you and your twin?" where the answer was "As alike as two peas in a pod" or "Ordinary sibling likeness (like sisters or brothers)" was more predictive of the "true" zygoty of the twins compared to the overall score obtained from all of the questions from the PPQ. *The accuracy of the peas in a pod questionnaire in our study was found to be 92.5% and 97.2% for MZ and DZ twins respectively. This is very similar to results from Peeters' (1998) study which showed an accuracy of 92.8% and 97.1% for MZ and DZ twins respectively. The highly reproducible results give weight to the validity of the peas in a pod question in accurately diagnosing zygoty.*

This study has demonstrated that consistency of the response to the same set of questions, administered over a period of several years demonstrated even stronger predictability and that this response represents the "true" zygoty. When both of the twins in the pair consistently agreed that they were MZ or DZ, the responses were their "true" zygoty in 99.6% and 99.7% of twin pairs respectively. *95% of individuals who were categorised as UZ across longitudinal PPQs, or whose*

category changed across time, were DZ when genotyped. There are a number of limitations and biases present within this study. The study may not be generalizable to other twin populations as there is a high proportion of females within the TwinsUK cohort and a relatively older age of mean 61 years compared to other cohorts. *The TwinsUK Cohort (Moayyeri 2013) is predominantly female (~80%) as historically the initial focus of the study was osteoporosis and so middle-aged female twin pairs were recruited from 1992; despite subsequent inclusion of men, like many twin cohorts there is a female volunteer bias.*

They were born in the United Kingdom and so represent the cultural norms of their society and knowledge about twinning, and may be biased in that they are volunteers in a research cohort.

A further bias might occur through the fact that the questionnaire was asked on five separate occasions. Comparisons between the first time answering the questionnaire and the overall result after answering multiple questionnaires have shown that repeating the PPQ is more accurate than asking on a single occasion, but there is however the risk that the twins “learnt” their true zygoty and therefore the answers for questionnaires completed after this subsequently are a more accurate representation of their “true” zygoty.

We have been able to compare the zygoty determined by the PPQ questionnaire with the zygoty determined via (GWAS) genotyping in a large number of twins (n=3,859), but a possible limitation is due to the fact that only one twin was sent for genotyping from twin pairs that were thought to be MZ and both twins were sent for genotyping from twin pairs that were thought to be DZ. This means that there are fewer MZ twin pairs with genotyping data, *resulting in a greater number of DZ pairs than MZ pairs available for comparison with the PPQ zygoty data. Despite this we still had genotypic data on over 340 MZ twin pairs who self-reported as MZ.*

Using data from adult twins in the TwinsUK cohort, we have validated the PPQ as an excellent proxy indicator of zygoty. In particular, if an initial PPQ from both twins scores as DZ, they are 98.7% likely to be DZ on genotyping (and 99.1% likely if one scores DZ and the other UZ). While only 88%

of pairs where both initially scored MZ were truly MZ, this improved to 99.6% of pairs where they consistently scored MZ on serial PPQs. The single “alike as two peas in the pod” question was most discriminatory (92% and 97% accurate for MZ and DZ individuals). We would recommend that twin registries could use the PPQ as a quick and relatively inexpensive way of determining the zygoty of twins at registration. *It may be unnecessary to genotype pairs where both twins score as DZ on the PPQ (or one as DZ and one as UZ), and similarly where both twins have MZ scores over serial questionnaires.* However, depending on budget and time pressures, genotyping may be required to identify the true zygoty where there are other twin combinations (such as one twin scoring as MZ, or both UZ).

Acknowledgements:

KJW was funded by the ERC funded “EpiTwin” project (European Research Council (ERC 250157)) assessed the zygoty of the twins via the PPQ, compared the self-reported zygoty and the PPQ zygoty with the zygoty from the genotyping data and wrote the initial manuscript. ZJ made subsequent revisions in response to reviewers’ comments. MM is a Bioinformatician and Honorary Lecturer who is funded by the British Research Council and assessed the zygoty of the twins via the genotyping data and assisted in the writing and editing of the manuscript. LC adapted the original zygoty questionnaire by Nick Martin et al and acted as an advisor for this manuscript. RG, IGN and DY are funded by Wellcome Trust and advised on and edited the manuscript. TDS is an NIHR senior Investigator and is holder of an ERC Advanced Principal Investigator award and edited the manuscript, as did CJH who jointly directed the work.

Financial support:

This research received no specific grant from any funding agency, commercial or non-profit sectors.

Conflict of interest:

None

Ethical standards:

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. Written, informed consent was obtained from participants of the study.

References:

- Bønnelykke, B., Hauge, M., Holm, N., Kristoffersen, K., Gurtler, H. (1988). "Evaluation of zygoty diagnosis in twin pairs below age seven by means of a mailed questionnaire." *Acta geneticae medicae et gemellologiae* **38**(3-4): 305-313.
- Cederlöf, R., Friberg, L., Jonsson, E., Kaij, L. (1961). "Studies on similarity diagnosis in twins with the aid of mailed questionnaires." *Human Heredity* **11**(4): 338-362.
- Chen, J., Li, X., Chen, Z., Yang, X., Zhang, J., Duan, Q., & Ge, X. (2010). Optimization of Zygoty Determination by Questionnaire and DNA Genotyping in Chinese Adolescent Twins. *Twin Research and Human Genetics*, *13*(2), 194-200. doi:10.1375/twin.13.2.194
- Eaves, L. J., Silberg, J.L., Hewitt, J.K., Rutter, M., Meyer, J.M., Neale, M.C., Pickles, A.(1993). "Analyzing twin resemblance in multisymptom data: genetic applications of a latent class model for symptoms of conduct disorder in juvenile boys." *Behavior Genetics* **23**(1): 5-19.
- Heath, A. C., Nyholt, D.R., Neuman, R., Madden, P.A., Bucholz, K.K., Todd, R.D., Nelson, E.C., Montgomery, G.W., Martin, N.G. (2003). "Zygoty diagnosis in the absence of genotypic data: an approach using latent class analysis." *Twin Research* **6**(01): 22-26.
- Kasriel, J., Eaves, L. (1976). "The zygoty of twins: further evidence on the agreement between diagnosis by blood groups and written questionnaires." *Journal of Biosocial Science* **8**(03): 263-266.
- Lykken, D. T. (1978). "The diagnosis of zygoty in twins." *Behavior Genetics* **8**(5): 437-473.
- Magnus, P., Berg, K., Nance, W.E.(1983). "Predicting zygoty in Norwegian twin pairs born 1915–1960." *Clinical Genetics* **24**(2): 103-112.
- Moayyeri, A., Hammond, C.J., Hart, D.J., Spector, T.D. (2013). "The UK adult twin registry (TwinsUK Resource)." *Twin Research and Human Genetics* **16**(01): 144-149.
- Moayyeri, A., Hammond, C.J., Valdes, A.M., Spector, T.D.(2013). "Cohort profile: TwinsUK and healthy ageing twin study." *International Journal of Epidemiology* **42**(1): 76-85.
- Ooki, S., Yamada, K., Asaka, A., Hayakawa, K.(1989). "Zygoty diagnosis of twins by questionnaire." *Acta Geneticae Medicae et Gemellologiae* **39**(1): 109-115.
- Ooki S, Yokoyama, Y., Asaka A (2004). "Zygoty misclassification of twins at birth in Japan." *Twin Research* **7**(3): 228-232.
- Peeters, H., Van Gestel, S., Vlietinck, R., Derom, C., Derom, R. (1998). "Validation of a telephone zygoty questionnaire in twins of known zygoty." *Behavior Genetics* **28**(3): 159-163.

Purcell S, N. B., Todd-Brown K, Thomas L, Ferreira MZ, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007). "PLINK:a tool set for whole-genome association and population-based linkage analyses." *American Journal of Human Genetics*: 559-575.

Sarna, S., Kaprio, J., Sistonen, P., Koskenvuo, M. (1978). "Diagnosis of twin zygoty by mailed questionnaire." *Human Heredity* **28**(4): 241-254.

Song Y.M., Lee D., Lee M.K., Lee K., Lee H.J., Hong E.J., Han B., Sung J. (2010). "Validity of the zygoty questionnaire and characteristics of zygoty-misdiagnosed twin pairs in the Health Twin Study of Korea." *Twin Research and Human Genetics*: 223-230.

Spector, T. D. & F. M. Williams (2006). "The UK adult twin registry (TwinsUK)." *Twin Research and Human Genetics* **9**(06): 899-906.

Tables:

Table 1: Table of terminology

Term	Description
Dizygotic (DZ)	Non-identical twins originating from different ova, sharing half their segregating genes across the genome
Genome-wide association study (GWAS)	A hypothesis-free study where common variants (polymorphisms) across the whole genome are genotyped and individually tested across many individuals for association with a disease or trait
Hardy-Weinberg principle	A model stating that allele/genotype frequencies in a population will remain unchanged through generations in the absence of other evolutionary influences. The Hardy-Weinberg equation is used to calculate the genetic variation of a population at equilibrium.
Identity by descent (IBD)	A term used to describe a matching segment of DNA shared by two or more people that has been inherited from a common ancestor.
Linkage disequilibrium (LD)	The non-random association of alleles at different loci in a population.
Minor allele frequency (MAF)	Frequency at which the second most common allele occurs in a given population.
Monozygotic (MZ)	Identical twins originating from the same ovum, sharing all the same segregating genes.
Single-nucleotide polymorphism (SNP)	A DNA sequence variation where a single nucleotide in the genome is different between members of a species or paired chromosomes in an individual.

Table 2: Comparison of zygoty determined by first PPQ and genotype data for twin pairs

First PPQ Answer	Number of pairs in PPQ	Number of pairs who are MZ in GWAS (%)	Number of pairs who are DZ in GWAS (%)
Both answer Monozygotic (MZ)	343	304 (88.6%)	39 (11.4%)
Both answer Dizygotic (DZ)	390	5 (1.3%)	385 (98.7%)
Both answer Unknown Zygoty (UZ)	703	37 (5.3%)	666 (94.7%)
1 answers MZ and 1 answers DZ	36	10 (27.8%)	26 (72.2%)
1 answers MZ and 1 answers UZ	148	60 (40.5%)	88 (59.5%)
1 answers DZ and 1 answers UZ	186	4 (2.2%)	182 (97.8%)

Table 3: Comparison of overall zygosity determined by multiple PPQ's and genotype data for twin pairs

Zygosity from PPQ	Number of pairs in PPQ	Number of pairs who are MZ in genotype data	Number of pairs who are DZ in genotype data
Both answer Monozygotic (MZ)	274	273 (99.6%)	1 (0.4%)
Both answer Dizygotic (DZ)	312	1 (0.3%)	311 (99.7%)
Both answer Unknown Zygosity (UZ)	901	64 (7.1%)	837 (92.9%)
1 answers MZ and 1 answers DZ	5	1 (20.0%)	4 (80.0%)
1 answers MZ and 1 answers UZ	90	79 (87.8%)	11 (12.2%)
1 answers DZ and 1 answers UZ	224	2 (0.9%)	222 (99.1%)

Figures and Legends:

Figure 1: Flow diagram of zygoty data available from self-report, PPQ and genotyping

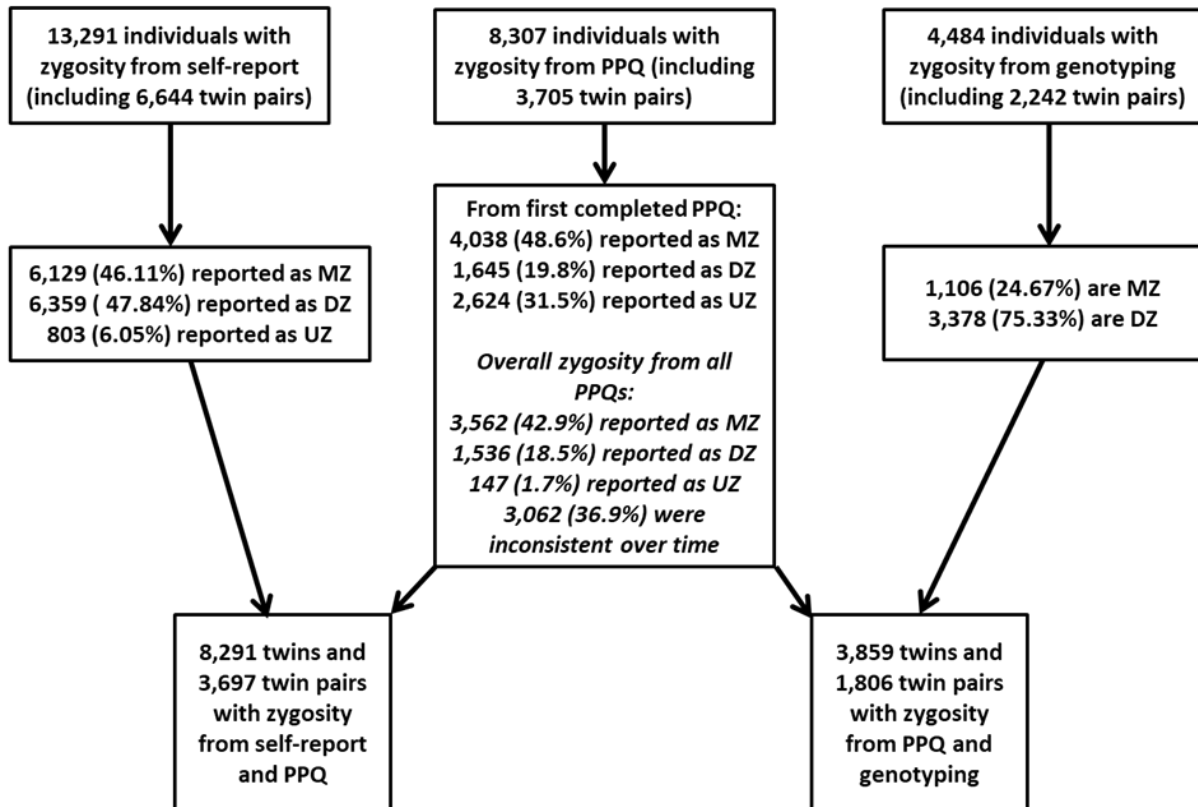


Figure 2: Percentage accuracy of the zygosity from the first PPQ with the zygosity from the genotyping data for individual twins

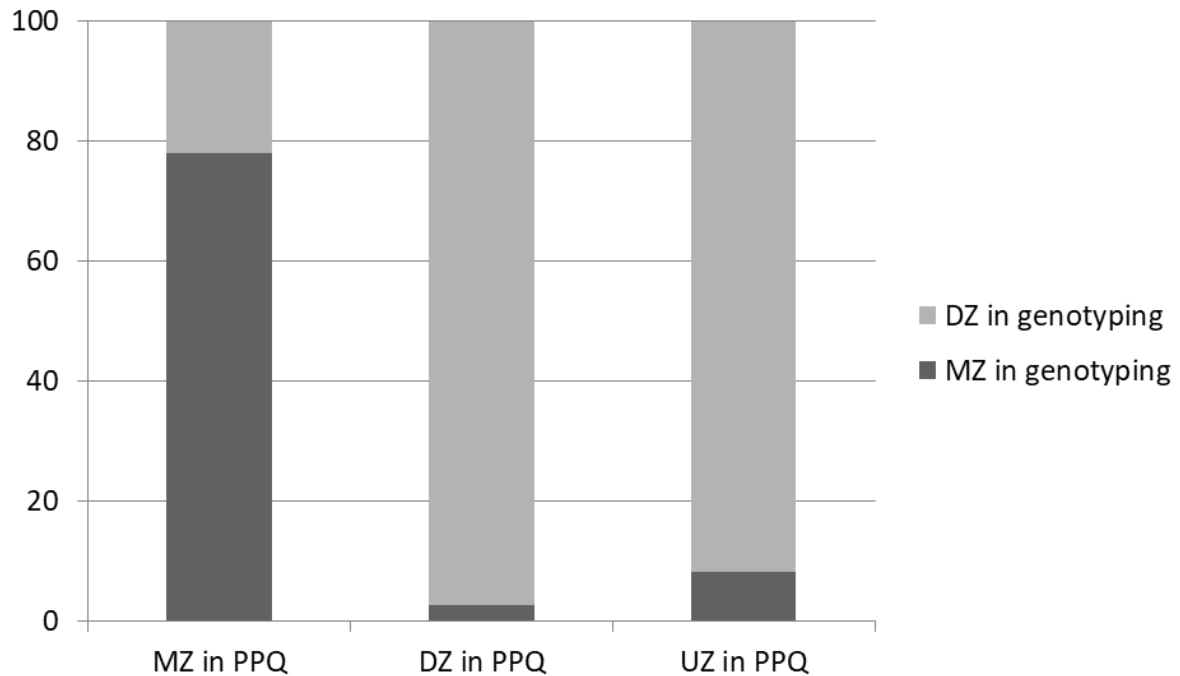


Figure 3: Percentage accuracy of the overall zygosity from all PPQ's with the zygosity from the genotyping data for individual twins

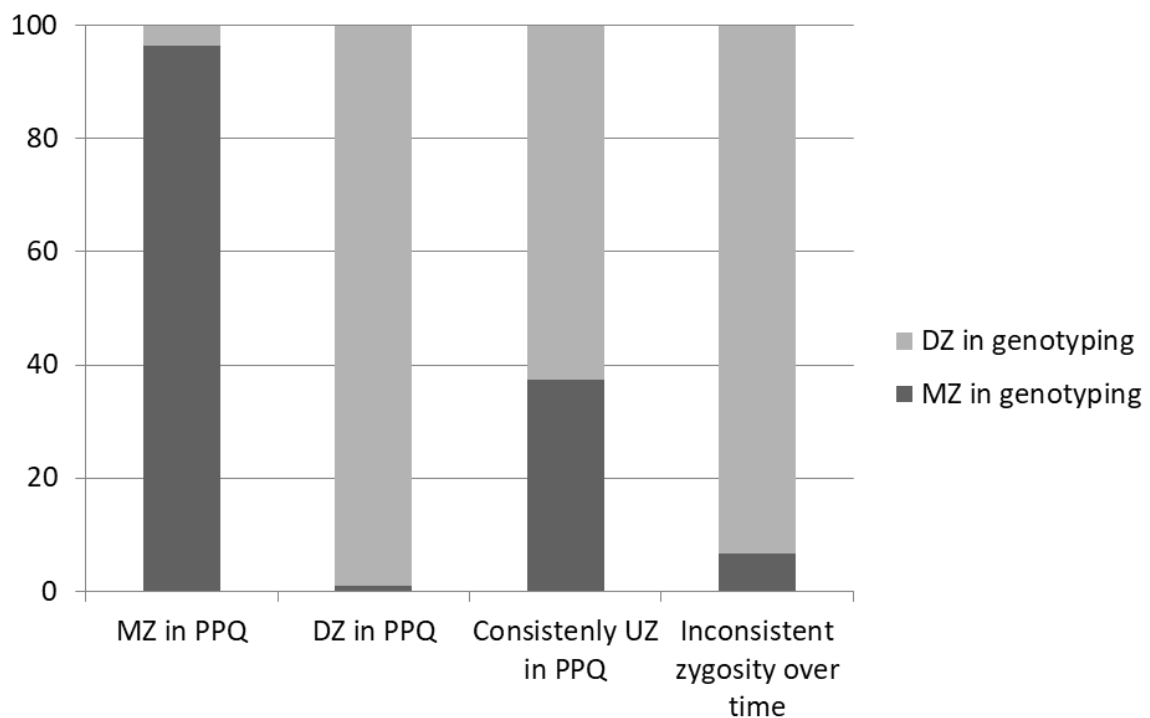


Figure 4: Percentage agreement between zygoty from PPQ and genotyping data for each question

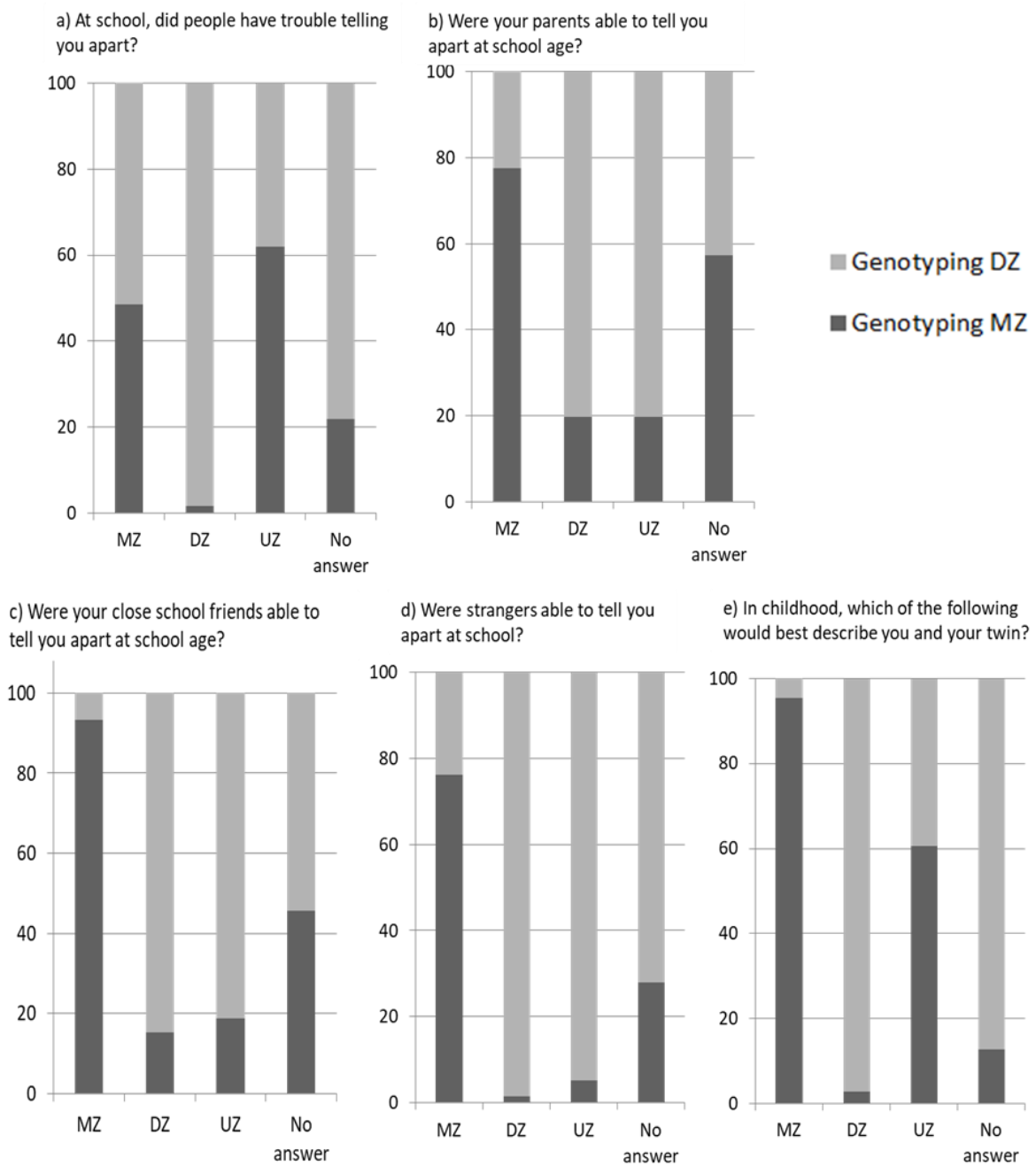
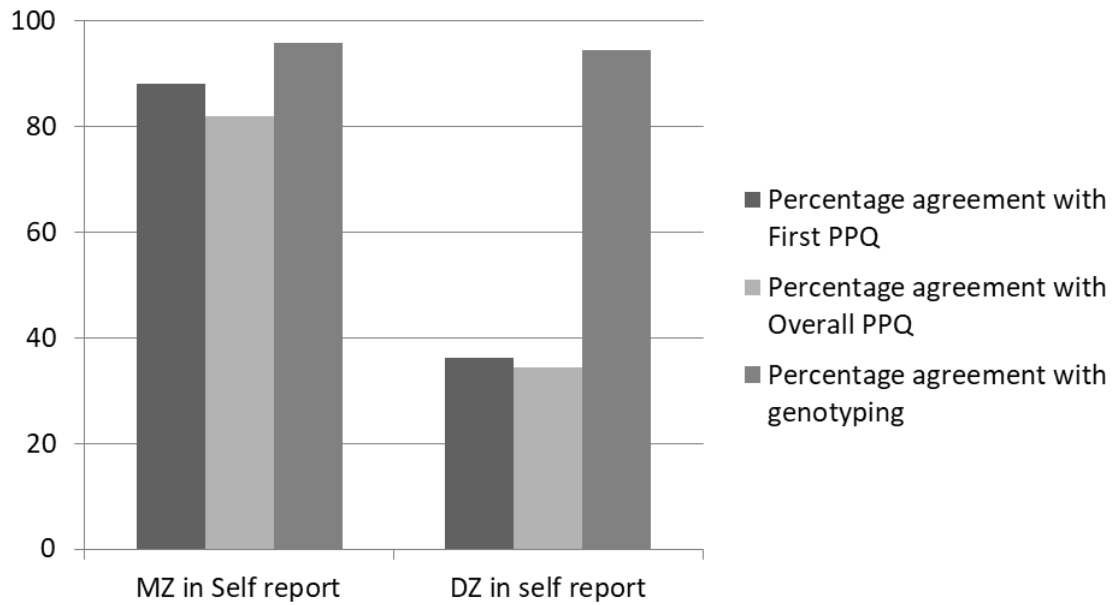


Figure 5: Percentage agreement between zygoty from self-report, first time answering the PPQ, Overall PPQ and genotyping data



Supplementary Material 1 – Peas in the Pod Questionnaire (PPQ)

ZYGOSITY

Dear Twin,

For research purposes we ask all twins to complete this short questionnaire so we can collect further information from which to ascertain your zygosity (i.e. whether you are an identical or non-identical twin), as some twins are uncertain about this. It is very important that you DO NOT confer with your twin otherwise it will give a false result. Please could you kindly complete the questions below and return with your questionnaire, even if you are sure about your zygosity.

Many thanks.

1. a) At school, did people have trouble telling you apart?

- (0) O Yes
(2) O No
(1) O I don't know

b) Were your parents able to tell you apart at school age?

- (2) O Yes
(0) O No
(1) O I don't know

c) Were your close school friends able to tell you apart at school age?

- (2) O Yes
(0) O No
(1) O I don't know

d) Were strangers able to tell you apart at school age?

- (2) O Yes
(0) O No
(1) O I don't know

e) In childhood, which of the following would best describe you and your twin? (Please select one)

- (0) O As alike as peas in a pod
(2) O Ordinary sibling likeness (like sisters or brothers)
(1) O I don't know

NOTE: We will contact you for further information if your Zygoty on completing this Questionnaire is different from your Self Report